AD________________

Grant Number DAMD17-96-1-6182

TITLE: Genes Altered by Intracisternal A Particles in Mouse Mammary Tumorigenesis

PRINCIPAL INVESTIGATOR: Michael R. Crowley

CONTRACTING ORGANIZATION: Health Research, Incorporated
Buffalo, New York 14263

REPORT DATE: August 1998

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
August 1998

3. REPORT TYPE AND DATES COVERED
Annual (1 Jul 97 - 30 Jun 98)

4. TITLE AND SUBTITLE
Genes Altered by Intracisternal A Particles in Mouse Mammary Tumorigenesis

5. FUNDING NUMBERS
DAMD17-96-1-6182

6. AUTHOR(S)
Michael R. Crowley

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Health Research, Incorporated
Buffalo, New York 14263

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

19981218 028

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200*

A mouse cDNA library constructed from a DMBA-induced primary mammary tumor which expressed high levels of intracisternal A particles was screened by differential hybridization using probes from the LTR and body of the element to find genes chimeric with IAPs. The 3' end of a gene, kokopelli, was isolated which showed tumor specific expression. The gene demonstrated tumor specific expression in all mouse mammary tumors and hyperplastic and neoplastic cell lines. Northern analysis of clones isolated from normal kidney, liver and brain mouse cDNA libraries suggested isolation of the normal counterpart to the tumor specific gene. Sequence analysis of these clones, however, revealed no significant homology outside a common B1 repetitive element. The cDNA clone from the kidney was localized to mouse chromosome 7 by genetic backcross analysis. Kokopelli was isolated as a chimera with U1 snRNP specific protein C. This association was further confirmed by isolation of 3' RACE products from a D2 tumor and from normal mammary gland. Further attempts to isolate the 5' end of the tumor specific transcript as well as the normal counterpart continue

14. SUBJECT TERMS Breast Cancer
Intracisternal A Particle, novel cancer gene, mammary chemical carcinogenesis, DMBA

15. NUMBER OF PAGES 30

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

MC Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

ae In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

ne In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

ne In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

ne In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

[signature] 7/30/98

PI - Signature Date

Table of Contents

INTRODUCTION

Endogenous retroviral (ERV) like sequences can represent agents of genetic variation within an organism through retrotransposition. Transcriptional activation of ERVs may lead to reintegration of the provirus into the host's genome causing mutations (1). These alterations may take the form of enhancer mutations affecting genes at sites distant from the point of integration, promoter mutations where genes are directly activated by the integration or disruption mutations where the retroviral sequences become incorporated into an adjacent gene. The disruption mutations may add sequences 5' through use of the LTR promoter, internally via splicing or 3' by donation of polyadenlyation signals (2). All of these chimeras can be assayed through differential screening with probes specific for elements within the retroviral element. Hybridization with probes specific for the *gag*, *pol* and *env* regions will identify the ERV transcripts contained within the given cDNA library. Rehybridization with probes specific for the LTR regions will discriminate between full length retroviral transcripts (represented by clones which hybridized to both sets of probes) and those which are potentially chimeric with LTR sequences (hybridized to only the LTR probes). This method of differential hybridization has been used successfully by Mager and colleagues to identify novel genes from an NTera2D1 cDNA library (2,3).

Previously it was reported that several DMBA (7,12-dimethylbenz (a) anthracene)-induced mouse mammary carcinomas over expressed the endogenous retrotransposon intracisternal A particles (IAPs). Indeed, Asch and Asch found that high expression of IAP RNA and protein is present in many mouse mammary tumors and preneoplasias, whereas little or no expression is detected in normal mammary glands from virgin, pregnant, lactating or involuting mice (4,5,6). Therefore changes in IAP expression frequently occur during the progression to tumorigenesis but not during normal growth and differentiation cycles of the mammary gland (5,6). For this reason a cDNA library from a DMBA-induced mouse mammary tumor which expressed the highest levels of IAP was constructed and screened by differential hybridization. Several clones were isolated which demonstrated chimerism between the IAP molecule and a unique cellular transcript. Many of these transcripts were found to be neither altered in expression patterns nor transcript size from

normal compared to tumor tissues upon Northern analysis. One clone however, showed both altered transcript size and tumor specific expression. The clone, p17b later renamed kokopelli, was chimeric with IAP sequences, however the association was deemed artifactual due to the presence of two polyadenylation signals and tails within the same clone. Sequence analysis of the isolated transcript demonstrated a murine B1 repetitive element in association with part of the U1 snRNP specific protein C gene separated by sequences not found in the database. Expression of the tumor specific transcript was documented in several mouse mammary tumors of various etiologies as well as in a series of mouse mammary cell lines which vary in their ability to grow tumors in nude mice. The transcript which did not show tumor specific expression was expressed in all mouse tissues and cell lines examined. This transcript is postulated to represent the normal counterpart to the tumor associated transcript.

Since only a partial clone was isolated in the original library screen, the full length tumor and normal transcripts were sought. Rescreening of the DMBA-induced tumor library resulted in transcripts which were unrelated to kokopelli or in isolation of the full length U1 snRNP specific protein C gene. Rapid amplification of cDNA ends was attempted but yielded no new transcript information as the products terminated within the B1 repetitive element of kokopelli. The efforts to isolate the transcript from normal tissues and to characterize the 3' end of the tumor transcript are the subjects of this report.

EXPERIMENTAL METHODS

cDNA Library Screening. The kidney (λZAP, Stratagene), brain and liver (λgt10, Clontech) cDNA libraries used for screening were generous gifts from Dr. Deborah Nagle of Millennium Pharmaceuticals. The kidney cDNA library was separated into 10 pools of approximately 1 x 10^5 clones per pool. These pools were screened with PCR primers p17bGSP4 and p17bGSP6 (Fig.1) and the positive pools were subdivided into pools of approximately 1 x 10^4 clones per pool. These pools were screened by PCR as described above, the positive pools were plated onto NZCYM by standard protocols (7). The phage were transferred to Nylon membranes (MSI, Westboro MA), the filters were divided into 4 quadrants and the phage eluted off the membrane in SM media. The

eluted phage were screened by PCR as described above. The resulting positive pools were then plated and screened via hybridization using an amplicon from primers p17bGSP4 and p17GSP6 by standard methods. The brain and liver cDNA libraries were screened by hybridization using a 660bp HaeIII probe derived from 3' end of clone pK823 (see below).

cDNA Cloning and Sequencing. All clones derived from the kidney cDNA library were excised from the phagemid as described by the manufacturer (Stratagene). The lambda clones isolated from the mouse liver and brain specific libraries were plaque purified and the DNA isolated via liquid lysate (8). Inserts were liberated from the λgt10 vectors by EcoRI digestion. The inserts were purified from a 1% agarose gel using Geneclean (Bio 101) and ligated into the EcoRI site of pGEM7Z(+). Sequence analysis was performed in Roswell Park's core facility using the ABI automated DNA sequencer. The M13 universal forward and reverse primers were used in all sequencing reactions along with gene specific primers where indicated. All sequences were analyzed using either the FASTA and BESTFIT algorithms of the GCG program maintained by the University of Wisconsin at Madison or by using the BLAST algorithm at the NCBI.

Rapid Amplification of cDNA Ends (RACE). The Marathon-Ready kidney cDNA (Clontech) was used to generate 5' extensions of clone pK823 (see Results). Primers used for 5'RACE were: K823-1 (5'CCCCAGGTGGAGATTTGTCTAC3'), K823-2 (5'CACGCTGTATGATCTCCCGGAG3') 494 bp and 296 bp from the 5' end of pK823, respectively and K823-3 (5' CCCACTATAAATATACAGCTCCATGGGCTTC3'), K823-4 (5' TCCCCTGTCTGTTCACCCAGG3'), 275 bp and 140 bp from the 5' end of clone pK823-2, respectively. A hot start reaction was utilized in every case. Two rounds of PCR were done for each 5' extension. The 5' extension of clone pK823 was conducted with primer K823-1 consisting of 35 cycles of denaturation at 94°C for 45 sec, annealing at 62°C for 45 sec and extension at 72°C for 3 min followed by a nested reaction with primer K823-2 using the same PCR conditions. The initial 5' extension of clone pK823-2 was done with primer K823-3 using the following conditions: 94°C for 30 sec, 70°C for 5 min for 10 cycles then 94°C for 30 sec, 67°C for 5 min for an additional 25 cycles. The high annealing temperature of the first 10 cycles allows for

gene specific primer binding but not for the upstream anchor primer. A nested round of PCR was done with primer K823-4 using 94°C for 45 sec, 65°C for 30 sec and 72°C for 3 min for 30 cycles. All RACE PCR products were excised from a 1.5% agarose gel, genecleaned and cloned into pGEM T/A Easy (Promega). To facilitate sequence analysis of the K823 RACE products several subclones were generated by digestion with various restriction enzymes followed by religation of the vector.

For 3' RACE total RNA from a D2 tumor and normal mammary gland from a virgin animal was extracted from fresh frozen tissue using the TRI reagent (Molecular Research Center, Inc.). The RNA was poly $(A)^+$ selected using oligo dT column chromatography as described by the supplier (Molecular Research Center, Inc.). One to 3 micrograms of poly $(A)^+$ RNA was reverse transcribed with Superscript II RT (Gibco, BRL) at 45°C using the 3' RACE oligo (5'GGCCTAGGCCTTAAGGGCCCTAC$(T)_{12}$3') as described by the manufacturer. The initial PCR was done using p17bGSP4 as the upstream gene specific primer in conjunctions with the 3' RACE primer (5'GGCCTAGGCCTTAAGGGCCCTAC3'). PCR amplification consisted of a hot start reaction, 95°C for 5 min, 80°C hold where buffer, polymerase and $MgCl_2$ were added, followed by 25 cycles of 94°C for 45 sec, 58°C for 45 sec and 72°C for 3 min. A nested round of PCR using primer p17bGSP5R as the upstream gene specific primer was done under the same conditions. PCR was done with the High Fidelity PCR kit from Boehringer Mannheim (Indianapolis, IN). Amplification products were run on a 1.5% agarose gel and cloned as described above.

Northern Analysis. Total RNA (15 μg) or poly A+ (1-5 μg) from several murine tumors and normal tissues were electrophoresed through a 1.2% 2.2M Formaldehyde gel and transferred to Nylon membranes by standard protocols. Northern blots were either cut into strips (stripblots) or remained intact for hybridization. Probes for hybridization were generated by random priming (9). Hybridizations were done in 50mM $NaPO_4$, 1% BSA, 7% SDS and 1mM EDTA at 65°C for 16 to 20 hours. Blots were rinsed in 2X SSC then washed stringently in 0.2X SSC and 0.1% SDS at

58°C for 1 hour. Filters were subjected to autoradiography at -80°C under an intensifying screen from 2 days to 2 weeks depending on the probe.

Genomic Localization of pK823. A C57BL/6 X Mus spretus backcross panel from the Jackson Labs (Bar Harbor, ME) was screened via hybridization with an 800bp Hae III fragment of pK823. This probe represents the extreme 3'end of the know sequence and does not contain any repetitive elements. Hybridizations were done on the backcross panel digested with Pst I (a generous gift from Dr. Rosemary Elliott, Roswell Park Cancer Institute). The mapping data were analyzed using Map Manager v2.0 (Ken Manly, Roswell Park Cancer Institute).

RESULTS and DISCUSSION

Isolation of pK823. A PCR based method of screening a λZAP mouse kidney cDNA library was employed to isolate the normal counterpart to the tumor specific transcript p17b (Kokopelli). Screening of library pools with primers specific for Kokopelli followed by filter hybridization with a probe derived from the primers used to screen the pools, produced several positive clones. One of these clones, pK823, contained a 1.9kb insert which hybridized to a 6.6 knt transcript in mouse heart and a DMBA-induced tumor along with the 1.4 knt transcript found in the DMBA-induced tumor on a Northern strip blot (Fig. 2). The probe did not hybridze to the 6.6 knt transcript from total mouse liver RNA but a faint hybridization signal was detected to the 1.4 knt transcript (Fig. 2, center). pK823 did not hybridize to the 1.4 knt transcript from total mouse heart RNA (Fig. 2, center). Upon sequence analysis of the 5' end of pK823 a murine B1 element homology region was discovered (Figs. 2 and 3). Figure 3 shows 673 bp of sequence from pK823 and the B1 element homology region is undelined. This homology region covers 151 bp of pK823 and is 87% identicle to the mouse B1 repetitive element found in Genbank (Fig. 3B). Sequences 3' to the B1 repetitive element had no homology to any sequences within the NCBI database. The unique portions of pK823 had no sequence homology with Kokopelli outside of the B1 element (data not shown). To determine if the B1 element was responsible for the hybridization signals seen on the Northern strip blots, pK823 was digested with Hae III which resulted in fragment sizes of approximately 660 bp, 405 bp, 340 bp, 107 bp, 42 bp, 40 bp, and 13 bp. The 660 bp fragment

represented the extreme 3' terminus and the 405 bp fragment represented an internal fragment, neither of which contained the B1 repetitive element. Both of these probes hybridized to the correct transcripts on Northern blots (Fig. 2). Both the 5' and 3' ends of pK823 hybridize strongly to the 6.6 knt transcript from kidney however the probes differ in hybridation to other tissues (Fig. 2 right and left panels). Also, the probes hybridize to the 1.4 knt transcript, previously found expressed only in tumors (Fig 2 and the previous report). These results suggest that the 1.4 knt transcript may not be expressed tumor specifically but rather upregulated in mouse mammary tumors. Upregulation may come at the exprense of the larger 6.6 knt transcript as Figure 2 right panel shows increased expression of the 1.4 knt transcript but little expression of the 6.6 knt transcsript in both the DMBA-induced and hormonally (Dim3) induced tumors. The RNA from a mammary gland isolated from a virgin mouse shows little hybridization to either band (Fig 2., right). The ethidium bromide staining of the gel prior to transfere show approximatly equal loads of all the RNA (data not shown). In addition, the results also suggest that the B1 element alone was not responsible for the Northern hybridization pattern seen despite the apparent lack of sequence similarity between pK823 and Kokopelli outside the B1 element. Repeated attempts to hybridized the B1 portion of the clone have not been successful (data not shown).

Genomic Localization of pK823. Kokopelli (p17b) was previously mapped to the proximal end of chromosome 17 (our unpublished results). In order to determine if pK823 was part of the same locus and thus represents an alternative product of that locus, the Jackson Labs (C57Bl/6 X Mus spretus) X Mus spretus backcross was analyzed using the 660 bp Hae III fragment of pK823. Backcross analysis using Map Manager localizes the pK823 clone to mouse chromosome 7 (Fig. 4). This result suggest that pK823 and Kokopelli are not the same gene but may instead be members of a larger gene family.

5' RACE of pK823. Although pK823 and Kokopelli do not reside on the same chromosome and therefore are not the same gene, Northern analysis suggests they recognize similar transcripts which are differentially expressed between normal tissues and tumors. It is possible that the sequence obtained to this point is entirely noncoding and the two genes are indeed members of a

larger family of which the B1 repetitive element is common at the 3' end but sequences downstream are unique. In other words, the 3' terminal exon of the gene family is marked by the presence of the B1 repetitive element although the exons themselves are not similar. Since substantial sequence 5' to the B1 element had not been isolated from clones pK823, pB311 or pL11-6 (see below), 5' RACE of the pK823 clone from kidney using the Marathon Ready cDNA kit (Clontech) was attempted (Fig. 5A). Screening the cDNA with primers K823-1 and K823-2 (see Methods) resulted in a 1.5 kb extension of pK823. Sequence analysis shows 100% similarity from K823-2 through the end of pK823 (Fig. 5B). This represents approximately 300 bp of sequence overlap. pK823-2 was digested with EcoRI and HincIII to isolate a probe fragment from which the B1 repetitive element had been removed (Fig. 5A). This probe hybridized to the 6,6 knt transcript in RNA from normal tissues, kidney and brain, but not from RNA isolated from tumors (Fig. 5C). This probe dected the 1.4 knt transcript from all tissues except brain (Fig. 5C). Therefore, with identical sequence overlap between the two clones and hybridization to the same transcripts on Northern blot analysis these results suggest that pK823-2 is a *bona fide* extension of clone pK823. The extension has been completely sequenced in both directions yet there is no apparent open reading frame which suggests that this clone may still be within the 3' untranslated region of the transcript.

A second round of 5' RACE with primers designed from the 5' end of pK823-2 (K823-3 and K823-4, see Methods) extended the pK823 sequence another 1.6 kb (Fig. 6A). Sequence analysis again shows 100% sequence similarity for the 117 bp of overlap between the 3' end of K823-4and the 5' end of pK823-2 (Fig. 6B). However when this clone (pK823-4) was used as a probe on Northern blots it did not hybridize to the same transcripts as pK823-2 or pK823. In fact pK823-4 hybridized to transcripts of approximately 1.8 knt and 1.9 knt in the tissues examined (Fig. 6C). This clone was not fully sequenced. Further attempts to isolate other 5' extensions have generated clones of various sizes but are identical to pK823-4 by restriction digest analysis (data not shown).

cDNA Library Screens. In an attempt to isolate longer clones from normal tissues, mouse brain and liver cDNA libraries were screened by filter hybridization with probes corresponding to either kokopelli or pK823. Two positive clones, pB311 from the brain cDNA library and pL11-6 from the liver cDNA library, were subjected to Northern strip blot analysis (Fig. 7A and data not shown). Both of these clones hybridized to the 6.6knt RNA transcript from normal and tumor tissues, and to the 1.4 knt transcript found only in the tumors (Fig. 7 and data not shown). Again the probe hybridized weakly to the 1.4 knt RNA transcript in normal tissues (Fig. 7A). Interestingly, clone pB311 hybridized to only the 1.4knt transcript in a D2 tumor (Fig. 7A). Hybridation was also seen to smaller RNA transcripts and probably represent hybridization of the B1 portion of the clone to expressed B1 elements or to the 7SL RNA from which B1 elementt are derived (Fig. 7A). Complete sequence analysis of pB311 and pL11-6 revealed the presence of a murine B1 repetitive element in each clone (Fig. 7B and C). The B1 repetitive element homology region from both pB311 (Fig. 7B) and pL11-6 (Fig. 7C) are underlined. Homology to the B1 repetitive element ranged from 85% over 164 bp for pB311 to 80% over 133 bp for pL11-6. There was no significant sequence homology between pB311 and pL11-6 or with each clone and Kokopelli and/or pK823 (data not shown). The B1 repetitive elements from all four clones were only 80-85% homologous. pB311 and pL11-6 were not analyzed further.

3' RACE for Kokopelli. In the previous report Kokopelli was found to be chimeric with two other genes, namely IAP and U1 snRNP specific protein C. The association with the IAP molecule was deemed a recombination artifact due to the presence of two poly $(A)^+$ tails and signals within the clone. The association with the U1 snRNP was, however, not fully investigated. Therefore, in order to isolate the true 3' end of the Kokopelli gene 3' RACE was undertaken using primers upstream of the unique region/snRNP breakpoint. A D2 tumor which expressed high levels of the 1.4 knt transcript yet very low to undetectable levels of the 6.6 knt transcript (see Fig. 7A) was subjected to reverse transcrition using a unique oligo dT adapter primer (see Methods). The resulting cDNA was then subjected to two rounds of nested PCR which generated an approximately 600 bp amplicon. Sequence analysis of the amplicon revealed the exact structure as

that found from the original clone (Fig. 8A). Comparison of the the two clones reveals that they were not 100% identical (Fig. 8A). These results suggest that the tumor associated transcript was indeed chimeric with the U1 snRNP specific protein C. Furthermore, the sequence from the clone isolated from the D2 tumor predicts an inframe association which alters the downstream amino acid sequence and hence the protein of U1 snRNP specific protein C (Fig 8B).

To determine if the same structural arrangement was seen in normal tissues, 3' RACE was conducted on poly $(A)^+$ selected RNA from a mammary gland from a virgin mouse exactly as described above. The amplicon derived from the normal mammary gland was of similar size as that isolated from tumors. Sequence analysis demonstrated the same structural arrangement as that seen previously (Fig. 8A). These data suggest that the 3' end of Kokopelli is chimeric with sequences from the snRNP specific protein C. The mouse gene for the U1 snRNP specific protein C has recently been cloned and it was determined that there are many copies of the gene in the genome (10). The sequence identified from the D2 tumor and mammary gland may indeed represent the correct sequence of the transcript as they were isolated using a high fidelity polymerase mixture not present in the original cloning of kokopelli. Taken together, these data would suggest that the association of the unique regions of Kokopelli and the snRNP described above is not artifactual but rather genuine and encodes for a protein with a unique function. However, the protein predicted starts from the nested primer used in the 3' RACE procedure and is rather tenuous as the entire coding region of kokopelli has not been isolated from any tissue to date. Only after the complete transcript has been isolated can predictions be made on possible protein translation and function. Attempts to isolate the 5' end of Kokopelli continue.

RECOMMENDATIONS ON THE S.O.W.

For the current period of the grant the focus was on isolation of the normal counterpart to the tumor associated transcript. It has been postulated that the 3' end of kokopelli, as determined by Northern analysis, is common between the normal and tumor associated transcripts (the 6.6 knt and 1.4 knt, respectively). This commonality suggests the alteration which has occurred in the tumor transcript was 5' to the known sequence. Therefore, by isolation of the full length normal transcript probes could be generated along the 6.6 knt to use on Northern blots to determine what sequences were contained within the tumor specific transcript. To this end several normal cDNA libraries were screened to isolate the full length normal transcript, as outlined in task 2 (months 12-36). Three different transcripts from 3 different libraries hybridized to the same transcripts via Northern analysis as kokopelli, however these clones shared no sequence homology outside a common murine B1 repetitive element. Further attempts by library screening and PCR techniques to isolate both the full length tumor associated transcript and the full length normal transcript are currently underway.

Task 2 also outlined functional assessment of the cloned gene. This task is difficult to accomplish until the full length transcripts have been identified. Sequence analysis of the 3' RACE products from the D2 tumor and from the normal mammary tissues demonstrated an open reading frame. Amino acid translation was carried out but no recognizable functional domains were detected (i.e. transactivation domains, DNA binding domains, transmemebrane domains, etc.). This analysis is tenuous due to the incomplete nature of the clone.

Task 3 of Southern analysis of normal and tumor tissue has been accomplished however due to hybridization problems the experiment has not been successful on the full panel of DMBA induced tumors. Also, determination of the clones role in oncogenic transformation will have to wait until the full length tumor associated transcript is isolated. Isolation of the genomic locus of the gene from normal and tumor tissue may, at this late date, be unaccomplishable. PAC and BAC mouse libraries are available for screening but emphasis should be placed on isolation of the cDNA before genomic clones are obtained.

Aspects of task 4 have been accomplished, screening of tumors arising from various etiologies and the DMBA-induced series of tumors for the occurrence of the tumor specific transcript. However, analysis of the mutants gene's role in development of DMBA-induced tumorigenesis may not be accomplished nor necessary. Experiments utilizing the TM series of mouse tumor cell lines has indicated that expression of the 1.4 knt maybe a very early event in tumor progression. The TM3L cells are hyperplastic and do not usually form tumors in nude mice yet the 1.4 knt transcript was expressed in those cells. In addition, the non-neoplastic cell line NOG-8 also demonstrated expression of the 1.4 knt transcript suggesting an early role for the gene in tumor formation.

CONCLUSIONS

Isolation of a novel gene fragment from a DMBA-induced mammary carcinoma has led to the possible identification of a large gene family with a unique genomic organization. Sequence analysis of transcripts isolated from atleast three different sources of RNA (DMBA-induced tumor, D2 tumor and normal mammary gland from a virgin mouse) demonstrate unique sequences chimeric and in frame with the U1 snRNP specific protein C. These sequences are expressed in all normal tissues examined as a 6.6 knt transcript while mammary tumors, neoplastic and hyperplastic cell lines express a unique 1.4 knt transcript. Clones isolated from normal mouse kidney cDNA library and the DMBA-induced tumor cDNA library hybridized to the 6.6 knt transcript on Northern analysis yet map to different regions of the genome by genetic backcross analysis supporting a multi gene hypothesis. However, only the original isolate, kokopelli, has the unique chimeric structure with the U1 snRNP specific protein C. All potential members of the gene family share a murine B1 repetitive element in their putative 3' untranslated regions.

REFERENCES

1). Kuff, E.L., and K.K. Lueders Ddv. Cancer Res. **51**, 183 (1988)

2). Feuchter, A.E., J.D. Freeman and D.L. Mager. Genomics **13**, 1237 (1992)

3). Feuchter-Murthy, A.E., J.D. Freeman and D.L. Mager. Nuc. Acids Res. **21**, 135 (1993)

4). Asch, B.B. and H.L. Asch. Cancer Res. **50**, 2404 (1990)

5). Asch, H.L., R.Z. Vinci, J.Y.Chang, F. Natoli, D.L. Stoler, G.R. Anderson and B.B. Asch. Cell Biol. Int. **17**, 961 (1993)

6). Asch, B.B., H.L. Asch, D.L. Stoler and G.R. Anderson. Int. J. Cancer **54**, 813 (1993)

7). Sambrook J., E. F. Fritsch, and T. Maniatis. Molecular Cloning: A Laboratory Manual. 2nd ed. The Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989)

8). Chisolm, D., BioTechnques **7**, 21 (1989)

9). Feinberg and B. Vogelstein. Anal. Biochem. **137**, 266 (1984)

10). Nelissen R. L., J. M. Gunnewiek, M. H. Lambermon, and W. J. Van Venrooij. Gene **184**, 273 (1997).

Figure1. Schematic representation of kokopelli. The mouse B1 repetitive element is shown as the dark striped box and is 256 bp in length. The U1 snRNP specific protein C is shown as the darkly stippled box. The lighly stipple box represents the PCR amplicon from primer GSP-4 and GSP-6 used as a probe on Northern blots and library screens. The PCR primers used 3' RACE are shown with direction or orientation. The light boxes are unique sequences.

Figure 2. Characterization of pK823. An approximate 1.9 kb clone was isolated from a normal mouse kidney library via PCR and hybridizations with kokopelli primers and probes, respectively. A partial restriction map is shown. Not all the Hae III sites are depicted for clarity. A 1.7 kb Eco RI fragment was isolated and used as a probe on Northern strip blots, center panel. The probe hybridized to an approximate 6.6 knt transcript in heart and tumor but not liver and to a 1.4 knt transript in the tumor lane only, center panel. Two sub-fragments of the clone, 405bp Hae III fragment and a 660 bp Hae III fragment, were used as probes on Northern strip blots to determine if the B1 repetitive element was responsible for the hybridization pattern seen with the full length probe. Both the 3' probe (660 bp Hae III) and the 5' probe (405 bp Hae III) hybridized to the same bands, a 6.6 knt in all tissues and a 1.4 knt in tumors, left and right panels. The 18S and 28S bands are marked. The B1 element homology region is shown as a dark striped box. The arrows above the clone are the direction and approximate length of sequencing. The entire clone has been sequenced in both directions, see Fig. 5.

Figure 3. Sequence comparison between pK823 and the mouse B1 repetitive element. A. The first 673 bp of pK823 are shown and the B1 repetitive element homology region is underlined. B. Sequence comparison using the Bestfit alogrithum in GCG of the B1 region of pK823 (top strand) and the mouse B1 repetitive element (bottom strand). There is 87% homology between the B1 elements covering 151 bp of pK823 sequence. The mouse B1 repetitive element can be found in Genbank under accession Gb_ro:Mmb1r.

Figure 4. Mapping of pK823. The (C57BL/6 X Mus spretus) X Mus spretus backcross panel from the Jackson labs digested with Pst I was a gift from Dr. Rosemary Elliott, Roswell Park Cancer Institute. The 660 bp Hae III fragment of pK823 (see Figure 2) was used as a probe against the backcross panel. The probe recognized a 1.5 kb allele from C57BL/6 and a 800 bp allele from Spretus. This polymorphism was followed in 94 backcross progeny. The top of the figure represents haplotype analysis of the middle of chromosome 7. The lower panel shows diagramtically the position of pK823 on chromosome 7 with distance in cM on the left side. The lod score for pK823 with D7Trk1 was 28.1 showing convincing linkage with this marker.

Figure 5. 5' RACE of pK823. A. The kidney Marathon RACE cDNA from Clontech was screened using primer K823-1 followed by a nested reaction with primer K823-2 (shown as arrows with 1 and 2 labeled above). Arrows above show direction and approximate length of sequence analysis of pK823-2 along with the sequence analysis of pK823. The B1 element is shown as a striped box for orientation. The partial restriction map is that proposed for the pK823/pK823-2 combination. Those restriction sites in parenthesis are part of the either the adaptor primer used in the RACE procedure or part of the vector. Again, not all the Hae III sites are shown for clarity. B. The sequence of the overlap region of the two clones with pK823 as the top strand and pK823-2 as the bottom strand. Note that the two are nealy 100% identicle. C. Northern strip blot with a 1.2 kb Eco RI/ Hinc II fragment from pK823-2 as a probe. The probe recognizes the 6.6 knt in all tissues as well as the 1.4 knt in the tumor. The chekered box represents pK823-2 and the horizontally striped box represents pK823. Primers K823-3 and K823-4 are shown at the extreme 5' end of the clone and were used to extend the RACE.

Figure 6. Isolation of pK823-4. Another round of 5' RACE using the kidney Marathon RACE kit from Clontech was used with primer K823-3 followed by a nested PCR with primer K823-4 (arrows 3 and 4 at the end of pK823-2, see also Figure 5). Again the striped box represents the B1 element for orientation. All three clones together are appoximately 4.6 kb. B. The region of

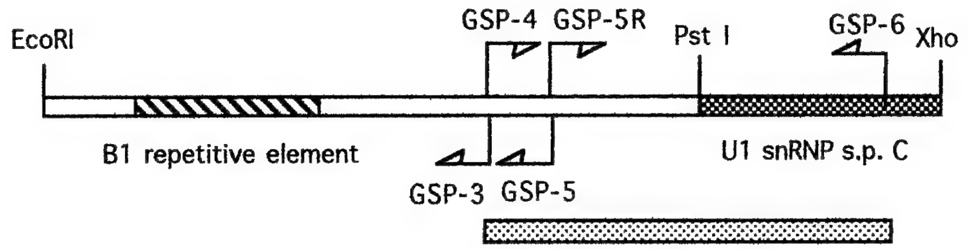
pK823-2 and pK823-4 which overlap. The 117 bp of overlap between the two clones is nearly 100%. C. Northern stip blot using the entire 1.8 kb pK823-4 clone as a probe. This clone hybridizes to two bands in the kidney, 1.8 knt and 1.9 knt, and only to the 1.9 knt transcript in other tissues. All sizes are relative to the 18S and 28S ribosomal bands which are marked. Lanes which are marked with an A+ are poly $(A)^+$ RNA and lanes without the A+ are total RNA from that particular tissue. Dim 3 is a hormonally induced mammary tumor and C3H is a MMTV induced mammary tumor from a BALB/c mouse foster nursed onto a fC3H mouse to introduce the mouse mammary tumor virus.

Figure 7. Sequence analysis of pB311 and pL11-6. A. The mouse brain cDNA isolate pB311 was used as a probe against Northern strip blots to determine expression patterens. The 800 bp Eco RI fragment hybridzed to 6.6 knt in brain, kidney and the TM10 mouse mammary cell line, and to the 1.4 knt transctript in the TM10 cell line as well as the D2 tumor. Weak hybridization to the 1.4 knt transcript could be detected in normal tissues with this probe. The pB311 probe did not hybridize to the 6.6 knt transcript in the D2 tumor. The lower hybridizing bands most likely represent cross hybridization to the B1 repetitive element portion of the probe. B. Sequence analysis of pB311. The B1 repetitive element homology region is underlined. C. Sequence analysis of pL11-6. The B1 repetitive element homology region is underlined. Neither of these two clones showed homology to known genes in Genbank.

Figure 8. Sequence comparison of kokopelli with the D2 tumor clone. A. The 3' RACE procedure was done on poly $(A)^+$ RNA from a D2 tumor and normal mammary gland from a virgin mouse using a unique oligo dT adaptor primer and GSP-4 followed by GSP-5R (see Methods). The D2 and normal mammary gland clones are shown as one sequence (D2/MG 3'RACE) as they were 100% identicle. Note at position 24 the presence of an extra C residue in the kokopelli transcript which is not prsent in the two other isolates. This extra C residue in kokopelli alters the open reading frame causing a stop codon downstream. Other changes are apparent between the

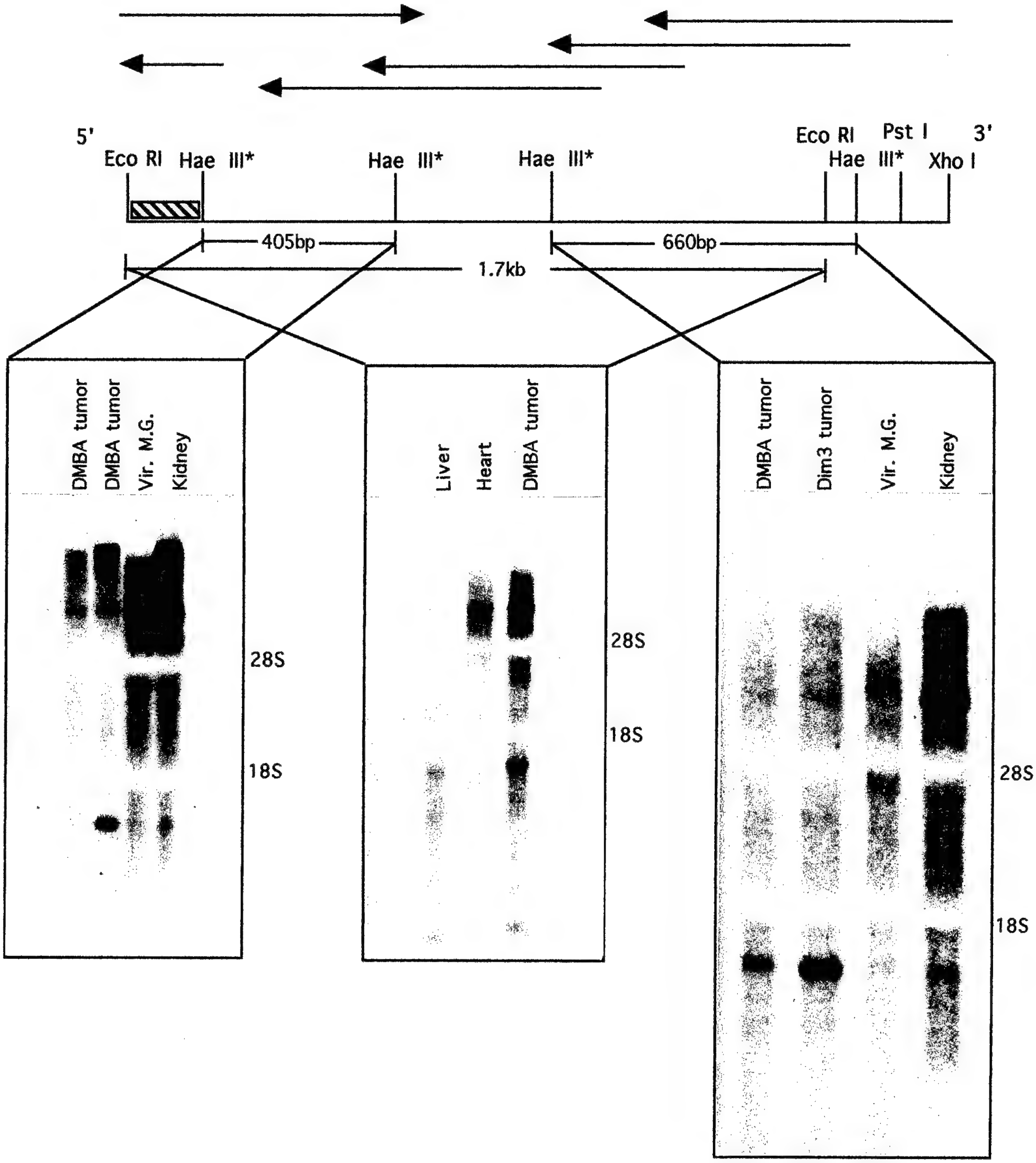
two clones, position 147 a C to T; an extra A residue in kokopelli at position 199; a TT to AA at position 377, 378. The nucleotide changes are shown in bold face type. B. The protein translation of the D2 and mammary gland transcripts using the single amino acid code.

Figure 1.



EcoRI
GSP-4
GSP-5R
Pst I
GSP-6
Xho I
B1 repetitive element
U1 snRNP s.p. C
GSP-3
GSP-5

Figure 2.



5'
3'
Eco RI
Hae III*
Hae III*
Hae III*
Eco RI
Pst I
Hae III*
Xho I
405bp
660bp
1.7kb
DMBA tumor
DMBA tumor
Vir. M.G.
Kidney
28S
18S
Liver
Heart
DMBA tumor
28S
18S
DMBA tumor
Dim3 tumor
Vir. M.G.
Kidney
28S
18S

Figure 3.

A.

```
  1  CCAATAGAGG GGATTTTTTT TTCTAATGTT GAAAAAGCTA CAAGCTACCA
 51  GGCAGTGGTG GCACATTTCT TTAGTCCCAG CACTTGGGAG GCAGAGGCAG
101  GTGGATTTCT GAGTTCTAGG CCAGCCTGGT CTACAGAGTA AGTTCCAGGA
151  CAGCCAGGGC TACACAGAGA AACCCTGTCT TGAGGGGGTG GGGGGGGTGC
201  AGAAAGGAAG GTAGGAAAGA AGGAAGGATC GAACTAAGAG CCAAAATACC
251  ACTCAATAAA ACTCCGGGAG ATCATACAGC GTGTGGGAGC CAAAGGAAAA
301  AAAATGAACA GTTTGGTTTT TAAAAATATT CTGTGGATGT AAAAAGGCAG
351  TGAGCTTAGA AACGATTTAT GAAGTCCCAT TTATTGAGTG GTTCCCTTAA
401  CTAGTCATCT CCCTGAGAAA ACAATGACCA TTCAAGGGAC ATTAGTATCT
451  TGAAGGTGTG TAGACAAATC TCCACCTGGG GATATCTCTT AACTCTCACT
501  ATTGCAGAAG GTGTTGAGGA GGCAGGCCTC GGGTGGGGGT TGGGAAGCAG
551  GCACTGTGTG GACGGGCCCT CCCCAGCAGG TTGATTATTA TTTTGGTGAA
601  GGTTTTTATT TTTCTTTACC TTCTGTAGGC TCCACAATGC CTGGATTTTA
651  TTGATATTGG TATTCTGGGC TCC
```

B.

```
pK823                33 AAAAGCTACAAGCTACCAGGCAGTGGTGGCACATTTCTTTAGTCCCAGCA 82
                        |||||| |  |   | | ||| |||||||| ||   ||||| ||||||||
Murine B1 element   236 AAAAGCCAAGAAAGAGCCGGC.GTGGTGGCGCACGCCTTTAATCCCAGCA 284
                                 .         .         .         .         .
pK823                83 CTTGGGAGGCAGAGGCAGGTGGATTTCTGAGTTCTAGGCCAGCCTGGTCT 132
                        || |||||||||||||||| |||||||||||||| |||||||||||||||
Murine B1 element   285 CTCGGGAGGCAGAGGCAGGCGGATTTCTGAGTTCGAGGCCAGCCTGGTCT 334
                                 .         .         .         .         .
pK823               133 ACAGAGTAAGTTCCAGGACAGCCAGGGCTACACAGAGAAACCCTGTCTTG 182
                        ||||||| |||||||| ||||||||||||||||||||||||||||||| |
Murine B1 element   335 ACAGAGTGAGTTCCAGAACAGCCAGGGCTACACAGAGAAACCCTGTCTCG 384

pK823               183 A
                        |
Murine B1 element   385 A
```

Figure 4

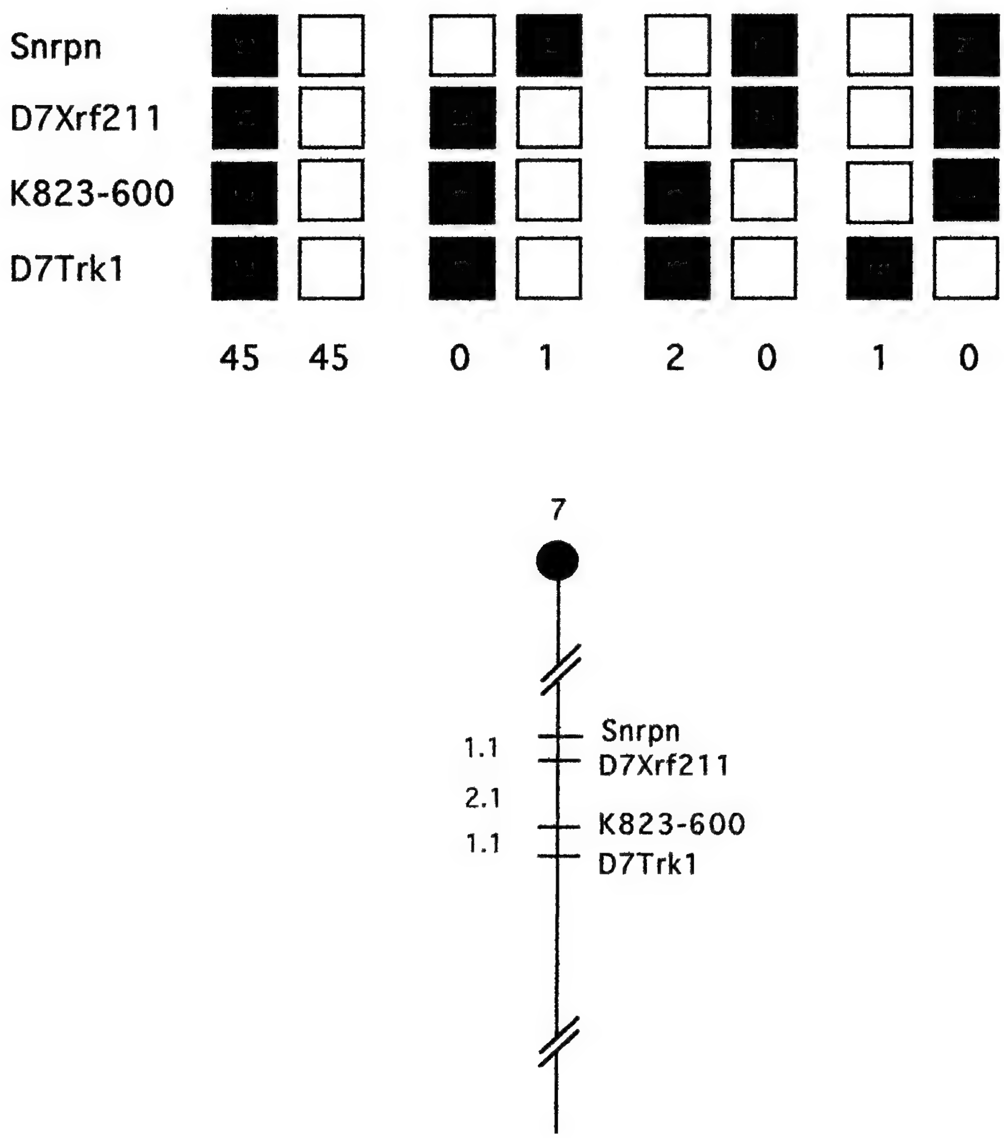
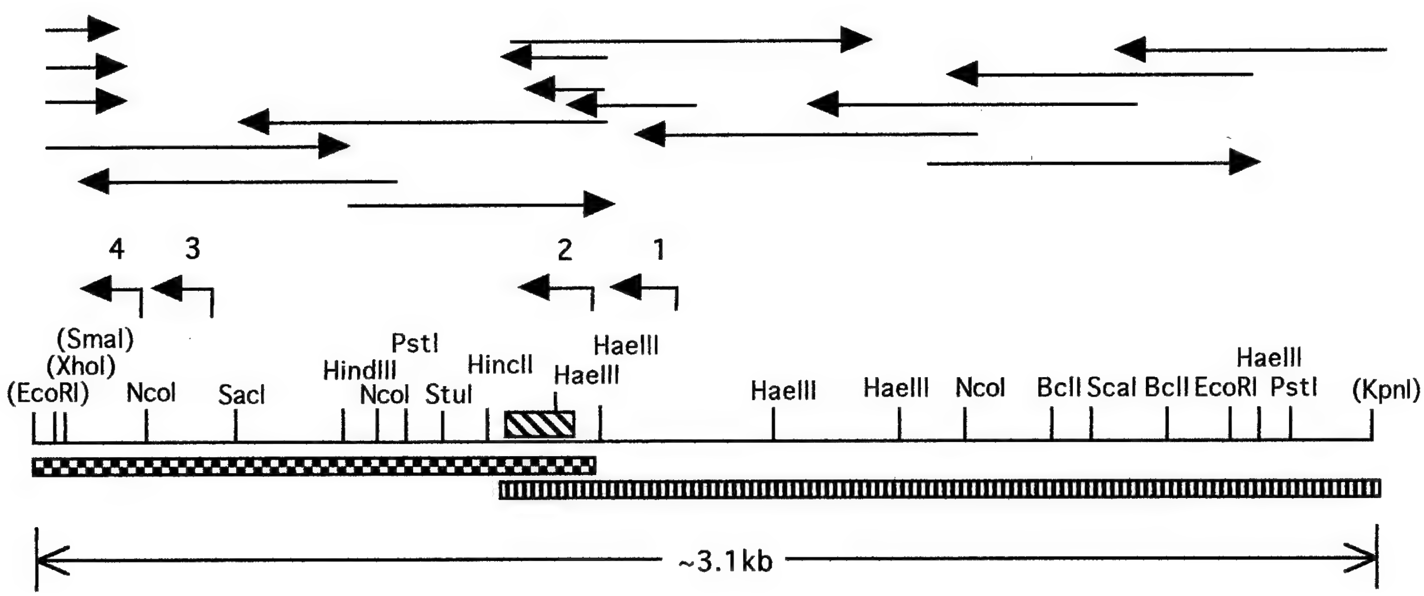

Snrpn
D7Xrf211
K823-600
D7Trk1
45
45
0
1
2
0
1
0
7
1.1
2.1
1.1
Snrpn
D7Xrf211
K823-600
D7Trk1

Figure 5.

A.



B.

```
K823    CCAATAGAGGGGATTTTTTTTTCTAATGTTGAAAAAGCTACAAGCTACCA
        ||||||||||||||||||||||||||||||||||||||||||||||||||
K823-2  CCAATAGAGGGGATTTTTTTTTCTAATGTTGAAAAAGCTACAAGCTACCA
                  .         .         .         .         .
K823    GGCAGTGGTGGCACATTTCTTTAGTCCCAGCACTTGGGAGGCAGAGGCAG
        ||||||||||||||||||||||||||||||||||||||||||||||||||
K823-2  GGCAGTGGTGGCACATTTCTTTAGTCCCAGCACTTGGGAGGCAGAGGCAG
                  .         .         .         .         .
K823    GTGGATTTCTGAGTTCTAGGCCAGCCTGGTCTACAGAGTAAGTTCCAGGA
        ||||||||||||||||||||||||||||||||||||||||||||||||||
K823-2  GTGGATTTCTGAGTTCTAGGCCAGCCTGGTCTACAGAGTAAGTTCCAGGA
                  .         .         .         .         .
K823    CAGCCAGGGCTACACAGAGAAACCCTGTCTTGAGGGGGTGGGGGGGGTGC
        ||||||||||||||||||||||||||||||||||||||||||||||||||
K823-2  CAGCCAGGGCTACACAGAGAAACCCTGTCTTGAGGGGGTGGGGGGGGTGC
                  .         .         .         .         .
K823    A..GAAAGGAAGGTAGGAAAGAAGGAAGGATCGAACTAAGAGCCAAAATA
        |  |||||||||||||||||||||||||||||||||||||||||||||||
K823-2  AGGGAAAGGAAGGTAGGAAAGAAGGAAGGATCGAACTAAGAGCCAAAATA
                  .
K823    CCACTCAATAAAA
        |||||||||||||
K823-2  CCACTCAATAAAA
```

C.

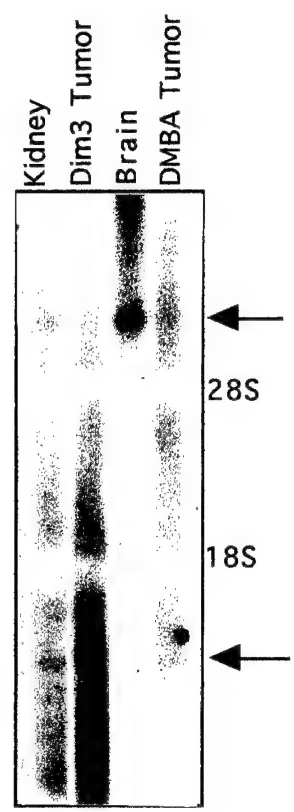
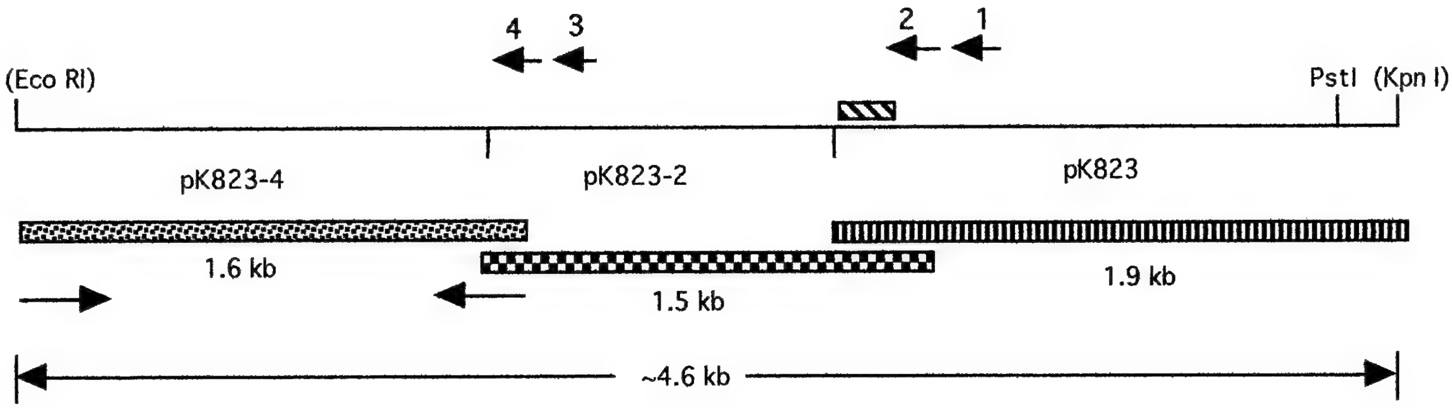


Figure 6.

A.



B.

```
K823-2   GCTCTTAGCAACATTTTCCATAGTATTTTCTCAAAATGGTGCGTTAGATA
         ||||||||||||||||||||||||||||||||||||||||||||||||||
K823-4   GCTCTTAGCAACATTTTCCATAGTATTTTCTCAAAATGGTGCGTTAGATA
                  .         .         .         .         .
K823-2   TAATCTAGGGGTTCCTCAGTATAAGCGCAAAACACACAGTAAGATACTAG
         ||||||| ||||||||||||||||||||||||||||||||||||||||||
K823-4   TAATCTA.GGGTTCCTCAGTATAAGCGCAAAACACACAGTAAGATACTAG
                  .
K823-2   CCAATGGTACATGTAGC
         ||||||||||||||||:
K823-4   CCAATGGTACATGTAGN
```

C.

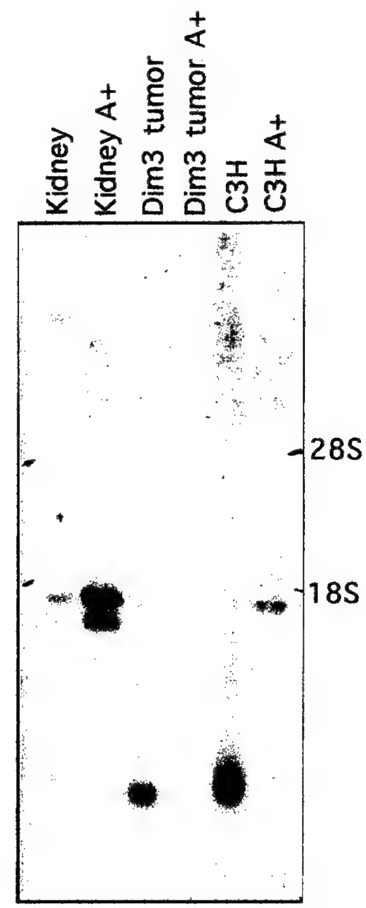
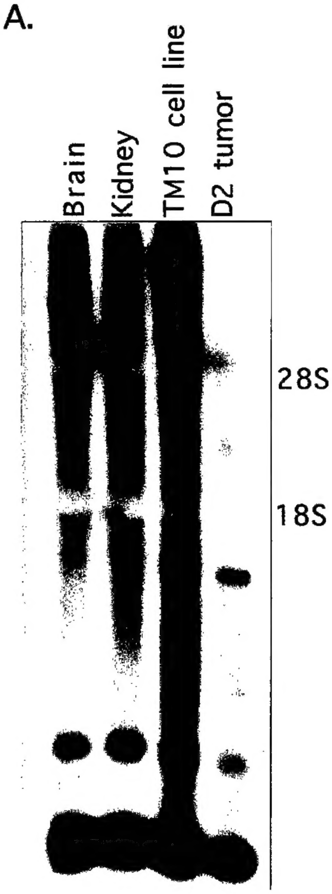


Figure 7.

A.



B.

1 TGGCAACATA GTGAGACCTG AGCTAAAATT AAAAATAAGG GGGGAAGAGG

51 CTTAGGAAGA TAGCTCAGTG GGTAGAGCAA ATGGGTTGAG CCTGAGTCAA

101 ACTCCCCAGC ACCCATCTAA AGAAGCTGGT AACCCAGCAC TGTGTGGGGC

151 AGATGCTGTC AGCTTGGCTC CCGATTCAGG GGAAAACCCA ATTTTCAGGA

201 ATAAGGAAGA GAGTGGCAGA GTGGAACACC AGATGGATGA ACGTCCTCTG

251 TTGGCCTCTG CTTGCGCTCA CAGGTGTGTG CATATTCACA CACATGCATA

301 CATCACACAC CAAAAGGAAA AAGGGAAGCA GAACGTCCCA TCTATCTCTC

351 TAGTTTCCCC AACAGTGTCC AACAGAGAGC AGAACAGCAA ATGTTACTTA

401 TTGTACGAAT TCTGTCAAAG AAAAGTCAAA GGAAAATTTC AAGCAGTTCT

451 AAATCTCCAG TACCAGCAAT AACTTGGATC TGAAGTCCTC TTTCAAAAGG

501 CTCGCTACAG CATTCTAAGG AGGTCAGCTC GTGCAGGTGG CAAGGGTCGT

551 CGTCGTCGTC GTTGTTGTAT TACACAGTAG GAAAGACTAA GCTTTTAGGG

601 GTGGTGGCAC AGAAGCCCTT AGACAATAGA GTACCAGCCG GGCGTGGTGG

651 CGCATGCCTT TAATCCCAGC ACTTGGGAGG CAGAGGCAGG CGGATTTCTG

701 AGTTCGAGGC CAGCCTGGTC TACAGAGTGA GTTCCAGGAC AGCCAGGGAT

751 ATACAGAGAA ACCCTGCCGG AATTCTTTTG CTTTTTACCC TGGAAGAAAT

801 ACTCATAAGC CACCTCTGTT A

Figure 7. continued

C.

```
  1 GAATTCGGCT GGCGTGGCAT TCTAAGTTTC TAGAAAGGCT CATGGAGACT
 51 GTTCAGCATT TCTTTGTCTA GAAGATAAGG GTAGGGATAA TGAGATTTTT
101 CTTGTGGAAT TAAGGAATGT ATTTTGCATA ACTCTTGTGA GCAGATGTTA
151 TATTAGTGTC TCTTGCTTCT CTGTCAGAGC CTGTGCAGCT GGCCTTGAAC
201 TTCAGGGGCT TTGTGATGCT TAGCCTACTC TATGTTGAAT ATGGTGCTGA
251 ACTGTGAAGT TTATTAGTTC TTACAATGTA AAAGAAACCC AGCTTGATGG
301 GCTGGAGAGA TGGCTCAGTA GTTAAGGGCA CTGACTGTCC TTCCAGAGGT
351 CCTGAGTTCA ATTCCCAGCA CCCACATGGT GGCTCACCAT GGGATCCCAT
401 GCCCTCTGAA GATAGCTACA ACATACTCAT ATAAATAATT TTTTTTTTAA
451 GAAAAAAAAC CCAGCTTGGG TACTATGTAA CAATATGCAC GCCTTTAGTC
501 CCAGCACTTG GGAAGCAGGT GGATCTTTTG GTTTAAGGCC AACCTGGTCT
551 ACAGGAGAGT TTTAGGATAG CCAGGGCTGC ACAGAGAGAT CCTGTCTCAA
601 GAAAAAAAGA CATTGCTATC CCTTTATCAT AGGGTTTTGG TTTTTTTTTT
651 TTTTGGGGGG GTGTTTTTGT TTTTTTTTTA ATTTTTCTGG GACAGGTTTC
701 TCTGTGTAGC CCTGACTGTC TCCCCTGTGC CTCCCAGGTG CTGGGATTAA
751 AGGCGTGCAC CACCACTGCC CGGCTTTTAT CATAGTTTTT AGTTGAAATT
801 ATTTTTTGGT ATGTCTGTAG CAGCTGGTTT GTTGTTTGTG TCCCTATGTG
851 GACCCACACT GGCCTCAGAC TGATGTTCCT GTCTCTCTCT CCCCAGTGCA
901 TAATGGCAAA TATTAAACTT AACGTTTTTT CAAAAAAAAA AAACCGGAAT
951 TC
```

Figure 8.

A.

```
Kokopelli        CATGTGTGTTGTCACCGTCCTCCCCCATTCCCCCCAAAAAAACCCAAACA
                 ||||||||||||||||||||| ||||||||||||||||||||||||||||
D2/MG 3' RACE    CATGTGTGTTGTCACCGTCCT.CCCCATTCCCCCCAAAAAAACCCAAACA

Kokopelli        AACAGAGTTTCCCTGTGTATAGCCCTGGCTGTCCTGGAACTTACTCTGTA
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
D2/MG 3' RACE    AACAGAGTTTCCCTGTGTATAGCCCTGGCTGTCCTGGAACTTACTCTGTA

Kokopelli        GACAAGGCTGGCCCAGAATGCAAAGATTTTGTGTTGAGTATTGGGATTAA
                 |||||||||||||||||||||||||||||||||||||||||||||| |||
D2/MG 3' RACE    GACAAGGCTGGCCCAGAATGCAAAGATTTTGTGTTGAGTATTGGGACTAA

Kokopelli        AGGACCCCCCCCCCTTTTTAGGTTTAATGGACGTCCCGTCTGCTGTCAAC
                 ||||||||||||||||||||||||||||||||||||||||||||||| ||
D2/MG 3' RACE    AGGACCCCCCCCCCTTTTTAGGTTTAATGGACGTCCCGTCTGCTGTC.AC

Kokopelli        CTGAATGTTGTTTCTCACTCCTTTCCCTTTGTTCTGTTCTGCAGCTCCTG
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
D2/MG 3' RACE    CTGAATGTTGTTTCTCACTCCTTTCCCTTTGTTCTGTTCTGCAGCTCCTG

Kokopelli        GGATGAGACCACCCATGGGAGGCCACATGCCCATGATGCCCGGACCTCCC
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
D2/MG 3' RACE    GGATGAGACCACCCATGGGAGGCCACATGCCCATGATGCCCGGACCTCCC

Kokopelli        ATGATGAGACCTCCTGCCCGCCCTATGATGGTGCCCACCCGGCCTGGCAT
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
D2/MG 3' RACE    ATGATGAGACCTCCTGCCCGCCCTATGATGGTGCCCACCCGGCCTGGCAT

Kokopelli        GACCCGGCCAGACAGATAAGAGCAGTTGCGCTCTTGATGGTTTTGTATTT
                 ||||||||||||||||||||||||| ||||||||||||||||||||||||
D2/MG 3' RACE    GACCCGGCCAGACAGATAAGAGCAGAAGCGCTCTTGATGGTTTTGTATTT

Kokopelli        CTTGTTCTGTTCCACCAGGAGCTCTTGGTGCTGAGCCCGAGTGTTTACTA
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
D2/MG 3' RACE    CTTGTTCTGTTCCACCAGGAGCTCTTGGTGCTGAGCCCGAGTGTTTACTA

Kokopelli        GATGCATGGAAAGGAAACTTCCCTTCCTAACTGAATATTTTTGGAGGGAG
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
D2/MG 3' RACE    GATGCATGGAAAGGAAACTTCCCTTCCTAACTGAATATTTTTGGAGGGAG

Kokopelli        AAATAATACAAAAAAGTGCAGTTTTCACTTATATTGTGAAATGTGAAAAT
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
D2/MG 3' RACE    AAATAATACAAAAAAGTGCAGTTTTCACTTATATTGTGAAATGTGAAAAT
```

Figure 8 B

```
    CATGTGTGTTGTCACCGTCCTCCCCATTCCCCCCAAAAAAACCCAAACAAACAGAGTTTC
  1 ---------+---------+---------+---------+---------+---------+ 60
    GTACACACAACAGTGGCAGGAGGGGTAAGGGGGGTTTTTTTGGGTTTGTTTGTCTCAAAG

    H  V  C  C  H  R  P  P  H  S  P  Q  K  N  P  N  K  Q  S  F   -

    CCTGTGTATAGCCCTGGCTGTCCTGGAACTTACTCTGTAGACAAGGCTGGCCCAGAATGC
 61 ---------+---------+---------+---------+---------+---------+ 120
    GGACACATATCGGGACCGACAGGACCTTGAATGAGACATCTGTTCCGACCGGGTCTTACG

    P  V  Y  S  P  G  C  P  G  T  Y  S  V  D  K  A  G  P  E  C   -

    AAAGATTTTGTGTTGAGTATTGGGACTAAAGGACCCCCCCCCCTTTTTAGGTTTAATGGA
121 ---------+---------+---------+---------+---------+---------+ 180
    TTTCTAAAACACAACTCATAACCCTGATTTCCTGGGGGGGGGGAAAAATCCAAATTACCT

    K  D  F  V  L  S  I  G  T  K  G  P  P  P  L  F  R  F  N  G   -

    CGTCCCGTCTGCTGTCACCTGAATGTTGTTTCTCACTCCTTTCCCTTTGTTCTGTTCTGC
181 ---------+---------+---------+---------+---------+---------+ 240
    GCAGGGCAGACGACAGTGGACTTACAACAAAGAGTGAGGAAAGGGAAACAAGACAAGACG

    R  P  V  C  C  H  L  N  V  V  S  H  S  F  P  F  V  L  F  C   -

    AGCTCCTGGGATGAGACCACCCATGGGAGGCCACATGCCCATGATGCCCGGACCTCCCAT
241 ---------+---------+---------+---------+---------+---------+ 300
    TCGAGGACCCTACTCTGGTGGGTACCCTCCGGTGTACGGGTACTACGGGCCTGGAGGGTA

    S  S  W  D  E  T  T  H  G  R  P  H  A  H  D  A  R  T  S  H   -

    GATGAGACCTCCTGCCCGCCCTATGATGGTGCCCACCCGGCCTGGCATGACCCGGCCAGA
301 ---------+---------+---------+---------+---------+---------+ 360
    CTACTCTGGAGGACGGGCGGGATACTACCACGGGTGGGCCGGACCGTACTGGGCCGGTCT

    D  E  T  S  C  P  P  Y  D  G  A  H  P  A  W  H  D  P  A  R   -

    CAGATAAGAGCAGAAGCGCTCTTGATGGTTTTGTATTTCTTGTTCTGTTCCACCAGGAGC
361 ---------+---------+---------+---------+---------+---------+ 420
    GTCTATTCTCGTCTTCGCGAGAACTACCAAAACATAAAGAACAAGACAAGGTGGTCCTCG

    Q  I  R  A  E  A  L  L  M  V  L  Y  F  L  F  C  S  T  R  S   -

    TCTTGGTGCTGAGCCCGAGTGTTTACTAGATGCATGGAAAGGAAACTTCCCTTCCTAACT
421 ---------+---------+---------+---------+---------+---------+ 480
    AGAACCACGACTCGGGCTCACAAATGATCTACGTACCTTTCCTTTGAAGGGAAGGATTGA

    S  W  C  *

    GAATATTTTTGGAGGGAGAAATAATACAAAAAAGTGCAGTTTTCACTTATATTGTGAAAT
481 ---------+---------+---------+---------+---------+---------+ 540
    CTTATAAAAACCTCCCTCTTTATTATGTTTTTTCACGTCAAAAGTGAATATAACACTTTA


    GTGAAAATAAAGTCATCAGCTCTTTTAGTTAAAAAAAAAAAAAAAAAAAAAA
541 ---------+---------+---------+---------+---------+-- 592
    CACTTTTATTTCAGTAGTCGAGAAAATCAATTTTTTTTTTTTTTTTTTTTTT
```